

S.
Liu



1600

RAW SEQUENCE LISTING

DATE: 04/16/2002

PATENT APPLICATION: US/09/684,215A

TIME: 09:31:03

Input Set : A:\-80-1.app

Output Set: N:\CRF3\04162002\I684215A.raw

**Does Not Comply
Corrected Diskette Needed**

3 <110> APPLICANT: Skeiky, Yasir
4 Guderian, Jeffrey
5 Corixa Corporation
7 <120> TITLE OF INVENTION: Methods of Using a Mycobacterium tuberculosis Coding
8 Sequence to Facilitate Stable and High Yield Expression
9 of Heterologous Proteins
11 <130> FILE REFERENCE: 014058-008010US
13 <140> CURRENT APPLICATION NUMBER: US 09/684,215A
14 <141> CURRENT FILING DATE: 2000-10-06
16 <150> PRIOR APPLICATION NUMBER: US 60/158,585
17 <151> PRIOR FILING DATE: 1999-10-07
19 <160> NUMBER OF SEQ ID NOS: 22
21 <170> SOFTWARE: PatentIn Ver. 2.1
23 <210> SEQ ID NO: 1
24 <211> LENGTH: 1872
25 <212> TYPE: DNA
26 <213> ORGANISM: Mycobacterium tuberculosis
28 <220> FEATURE:
29 <223> OTHER INFORMATION: 32 KD serine protease MTB32A
31 <220> FEATURE:
32 <221> NAME/KEY: CDS
33 <222> LOCATION: (89)..(1156)
34 <223> OTHER INFORMATION: MTB32A
36 <220> FEATURE:
37 <221> NAME/KEY: sig_peptide
38 <222> LOCATION: (89)..(184)
39 <223> OTHER INFORMATION: N-terminal hydrophobic secretory signal sequence
41 <220> FEATURE:
42 <221> NAME/KEY: mat_peptide
43 <222> LOCATION: (185)..(1153)
45 <400> SEQUENCE: 1
46 gactacgttg gtgtagaaaa atcctgccgc ccggaccctt aaggctggga caatttctga 60
48 tagctacccc gacacaggag gttacggg atg agc aat tcg cgc cgc cgc tca 112
49 Met Ser Asn Ser Arg Arg Arg Ser
50 -30 -25
52 ctc agg tgg tca tgg ttg ctg agc gtg ctg gct gcc gtc ggg ctg ggc 160
53 Leu Arg Trp Ser Trp Leu Leu Ser Val Leu Ala Ala Val Gly Leu Gly
54 -20 -15 -10
56 ctg gcc acg gcg ccg gcc cag gcg gcc ccg ccg gcc ttg tcg cag gac 208
57 Leu Ala Thr Ala Pro Ala Gln Ala Ala Pro Pro Ala Leu Ser Gln Asp
58 -5 -1 1 5
60 cgg ttc gcc gac ttc ccc gcg ctg ccc ctc gac ccg tcc gcg atg gtc 256
61 Arg Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val

RAW SEQUENCE LISTING

DATE: 04/16/2002

PATENT APPLICATION: US/09/684,215A

TIME: 09:31:03

Input Set : A:\-80-1.app

Output Set: N:\CRF3\04162002\I684215A.raw

62	10	15	20	
64	gcc	caa	gtg	ggg cca cag gtg gtc aac atc aac acc aaa ctg ggc tac 304
65	Ala	Gln	Val	Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr
66	25	30	35	40
68	aac	aac	gcc	gtg ggc gcc ggg acc ggc atc gtc atc gat ccc aac ggt 352
69	Asn	Asn	Ala	Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly
70		45	50	55
72	gtc	gtg	ctg	acc aac aac cac gtg atc gcg ggc gcc acc gac atc aat 400
73	Val	Val	Leu	Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn
74		60	65	70
76	gcg	ttc	agc	gtc ggc tcc ggc caa acc tac ggc gtc gat gtg gtc ggg 448
77	Ala	Phe	Ser	Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly
78		75	80	85
80	tat	gac	cgc	acc cag gat gtc gcg gtg ctg cag ctg cgc ggt gcc ggt 496
81	Tyr	Asp	Arg	Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly
82		90	95	100
84	ggc	ctg	ccg	tcg gcg gcg atc ggt ggc ggc gtc gcg gtt ggt gag ccc 544
85	Gly	Leu	Pro	Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro
86	105	110	115	120
88	gtc	gtc	gcg	atg ggc aac agc ggt ggg cag ggc gga acg ccc cgt gcg 592
89	Val	Val	Ala	Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala
90		125	130	135
92	gtg	cct	ggc	agg gtg gtc gcg ctc ggc caa acc gtg cag gcg tcg gat 640
93	Val	Pro	Gly	Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp
94		140	145	150
96	tcg	ctg	acc	ggt gcc gaa gag aca ttg aac ggg ttg atc cag ttc gat 688
97	Ser	Leu	Thr	Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp
98		155	160	165
100	gcc	gcg	atc	cag ccc ggt gat tcg ggc ggg ccc gtc gtc aac ggc cta 736
101	Ala	Ala	Ile	Gln Pro Gly Asp Ser Gly Gly Pro Val Val Asn Gly Leu
102		170	175	180
104	gga	cag	gtg	gtc ggt atg aac acg gcc gcg tcc gat aac ttc cag ctg 784
105	Gly	Gln	Val	Val Gly Met Asn Thr Ala Ala Ser Asp Asn Phe Gln Leu
106	185	190	195	200
108	tcc	cag	ggt	ggg cag gga ttc gcc att ccg atc ggg cag gcg atg gcg 832
109	Ser	Gln	Gly	Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala
110		205	210	215
112	atc	gcg	ggc	cag atc cga tcg ggt ggg ggg tca ccc acc gtt cat atc 880
113	Ile	Ala	Gly	Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile
114		220	225	230
116	ggg	cct	acc	gcc ttc ctc ggc ttg ggt gtt gtc gac aac aac ggc aac 928
117	Gly	Pro	Thr	Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn
118		235	240	245
120	ggc	gca	cga	gtc caa cgc gtg gtc ggg agc gct ccg gcg gca agt ctc 976
121	Gly	Ala	Arg	Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu
122		250	255	260
124	ggc	atc	tcc	acc ggc gac gtg atc acc gcg gtc gac ggc gct ccg atc 1024
125	Gly	Ile	Ser	Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile
126	265	270	275	280

RAW SEQUENCE LISTING

DATE: 04/16/2002

PATENT APPLICATION: US/09/684,215A

TIME: 09:31:03

Input Set : A:\-80-1.app

Output Set: N:\CRF3\04162002\I684215A.raw

```

128 aac tgc gcc acc gcg atg gcg gac gcg ctt aac ggg cat cat ccc ggt 1072
129 Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly
130          285          290          295
132 gac gtc atc tgc gtg acc tgg caa acc aag tgc ggc ggc acg cgt aca 1120
133 Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr
134          300          305          310
136 ggg aac gtg aca ttg gcc gag gga ccc ccg gcc tga tttcgtcgcg 1166
137 Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala
138          315          320
140 gataccaccc gccggccggc caattggatt ggcgccagcc gtgattgccg cgtgagcccc 1226
142 cgagttccgt ctcccggtgc cgtggcatcg tggagcaat gaacgaggca gaacacagcg 1286
144 tcgagcacc ccccggtgcg ggcagtcacg tcgaaggcgg tgtggtcgag catccggatg 1346
146 ccaaggactt cggcagcgcc gccgccctgc ccgcgcgaccc gacctggttt aagcacgcgg 1406
148 tcttctacga ggtgctggtc cgggcgttct tcgacgccag cgcggacggt tccggcgatc 1466
150 tgcgtggact catcgatcgc ctgcactacc tgcagtggct tggcatcgac tgcattcgtt 1526
152 tgccgcgctt ctacgactcg ccgctgcgcg acggcggtta cgacattcgc gacttctaca 1586
154 aggtgctgcc cgaattcggc accgtcgacg atttcgtcgc cctggtcgac gccgctcacc 1646
156 ggcgaggtat ccgcatcatc accgacctgg tgatgaatca cactcggag tcgcaccctc 1706
158 ggtttcagga gtcccgccgc gaccagacg gaccgtacgg tgactattac gtgtggagcg 1766
160 acaccagcga gcgtacacc gacgccgga tcatcttcgt cgacaccgaa gagtcgaact 1826
162 ggtcattcga tcctgtccgc cgacagttct actggcaccg attctt 1872
165 <210> SEQ ID NO: 2
166 <211> LENGTH: 355
167 <212> TYPE: PRT
168 <213> ORGANISM: Mycobacterium tuberculosis )<220> insert
169 <223> OTHER INFORMATION: 32 KD serine protease MTB32A
171 <400> SEQUENCE: 2
172 Met Ser Asn Ser Arg Arg Arg Ser Leu Arg Trp Ser Trp Leu Leu Ser
173 1 5 10 15
174 Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala
175 20 25 30
176 Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu
177 35 40 45
178 Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Gly Pro Gln Val Val
179 50 55 60
180 Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr
181 65 70 75 80
182 Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val
183 85 90 95
184 Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
185 100 105 110
186 Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
187 115 120 125
188 Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
189 130 135 140
190 Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly
191 145 150 155 160
192 Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu
193 165 170 175

```

RAW SEQUENCE LISTING

DATE: 04/16/2002

PATENT APPLICATION: US/09/684,215A

TIME: 09:31:03

Input Set : A:\-80-1.app

Output Set: N:\CRF3\04162002\I684215A.raw

```

194 Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr
195           180           185           190
196 Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser
197           195           200           205
198 Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr
199           210           215           220
200 Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe Ala
201 225           230           235           240
202 Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly
203           245           250           255
204 Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu
205           260           265           270
206 Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val Val
207           275           280           285
208 Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile
209           290           295           300
210 Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp
211 305           310           315           320
212 Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Thr Trp Gln
213           325           330           335
214 Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
215           340           345           350
216 Pro Pro Ala
217           355
220 <210> SEQ ID NO: 3
221 <211> LENGTH: 396
222 <212> TYPE: DNA
223 <213> ORGANISM: Mycobacterium tuberculosis
225 <220> FEATURE:
226 <223> OTHER INFORMATION: 14 KD C-terminal fragment of MTB32A Ra12
228 <220> FEATURE:
229 <221> NAME/KEY: CDS
230 <222> LOCATION: (1)..(396)
231 <223> OTHER INFORMATION: Ra12
233 <400> SEQUENCE: 3
234 acg gcc gcg tcc gat aac ttc cag ctg tcc cag ggt ggg cag gga ttc 48
235 Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
236 1 5 10 15
238 gcc att ccg atc ggg cag gcg atg gcg atc gcg ggc cag atc cga tcg 96
239 Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
240 20 25 30
242 ggt ggg ggg tca ccc acc gtt cat atc ggg cct acc gcc ttc ctc ggc 144
243 Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
244 35 40 45
246 ttg ggt gtt gtc gac aac aac ggc aac ggc gca cga gtc caa cgc gtg 192
247 Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
248 50 55 60
250 gtc ggg agc gct ccg gcg gca agt ctc ggc atc tcc acc ggc gac gtg 240
251 Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val

```